Title: STREAMS2: GIS Metadata for Land Use, Stream Chemistry & Biological Data

1. Purpose

The Purpose of this Task Order is to provide services to the U.S. Environmental Protection Agency's (EPA) National Center for Environmental Assessment in compiling, creating, and organizing Geographical Information System (GIS) metadata for Task 2.1B under the Safe and Sustainable Waters research "Correlating Land Use, Stream Chemistry, and & Biological Condition". The resulting metadata will follow the most current EPA Geospatial Metadata Technical Specification and will be compliant with Federal Geographic Data Committee (FGDC) standards.

The metadata will be developed for all of the different types of GIS data used in task 2.1B, which may include, but are not limited to shapefiles of vector data and file geodatabases. The deliverable of this task will be GIS data with complete metadata that includes a description of the data and details about the spatial data and the attribute data. Metadata on the attribute data may include methods used to collect the data and/or geoprocessing steps done to produce the data. The Geospatial Data Publishing Workflow Standard Operating Procedure will be followed so that GIS data and metadata meet the requirements to be published on the EPA GeoPlatform Online Environment.

2. Background

The initial GIS data of sample points and watershed polygons were created as part of the West Virginia Regional EMAP (Detenbeck et al. 2004, Detenbeck et al. 2005, Detenbeck and Cincotta 2008). The 2004 report and the 2005 and 2008 journal articles are sources of metadata for the sample points and watershed polygons. Additional GIS data were obtained on mining permit boundaries, urbanized areas, and the 1:24000 National Hydrography Data High (NHDH) resolution for West Virginia. The initial GIS data are listed in table 1, and these will be provided to the Contractor. The Sources and Points of Contact listed in table 1 are additional sources of metadata. Geoprocessing steps for the watershed polygons are shown in figure 1, which led to classifying the watershed polygons into one of four types. Geoprocessing was also done to combine the GIS data of sample points with flat files of water chemistry, physical habitat, and fish data. Appendix A describes those geoprocessing steps that merged the sample points shapefile with water chemistry, physical habitat, and fish data collected at those points, and it describes the criteria applied to select the 82 sample points used in the analysis. The final two GIS data sets of the 82 sample points and 82 watershed polygons obtained from the geoprocessing are listed in table 2. Variables obtained from those final two GIS data were used in a multivariate spatial data analysis.

3. Scope of Work

The Contractor shall conduct the following tasks for each of the GIS dataset in accomplishing the objective of this Task Order. The contractor shall use the EPA Metadata

Editor (EME v3.0) for ArcGIS 9.2/9.3 (https://edg.epa.gov/EME/Download.htm). The contractor shall consult the appropriate literature, web sites, or points of contact to obtain the necessary information to complete the metadata for each GIS data set. The contractor shall describe the SAS code written to merge data sets and geoprocessing steps done in ArcGIS to the GIS data. The output of this task order shall be GIS data and metadata associated with the 82 sample points and watershed polygons that are compliant with the most current EPA Geospatial Metadata Technical Specification and FGDC compliance and that meet requirements to be published on the EPA GeoPlatform Online Environment (http://intranet.epa.gov/gis/geopolicies.html).

Task 1. Obtain GIS Data, Geoprocessing Descriptions, and Metadata Sources and Create Draft Metadata Using EME (v3.0) to Meet EPA standards and FGDC Compliance as Tested by the EME Validate Tool for the 82 Sample Points and Watershed Polygons

The contractor shall review the GIS data and sources of metadata listed in the references, tables 1 and 2, and the geoprocessing descriptions in figure 1 and Appendix A so as to produce a draft metadata for the shapefiles of the 82 sample points and watershed polygons meeting EPA standards and being FGDC compliant. The draft metadata of those two GIS datasets shall be viewable in ArcGIS 9.3.1. Examples of metadata content include giving the units of measurements for the physical habitat and water chemistry variables measured at the sample points, fish species names collected at the sample points and their abundances, and providing citations of the report and journal articles describing the methods used to collect the data. Specific physical habitat variables requiring metadata include: thalweg mean depth (cm), mean wetted width (m), and mean embeddedness of the channel plus margins (%). Water chemistry variables requiring metadata include: concentrations (mg/L) of calcium (Ca^{2+}), chloride (Cl^{-}), manganese (Mn), and sulfate (SO_4^{2-}) as well specific conductance (μS/cm). Appendix B contains the fish species name and the concatenated form of the names used in the GIS point shapefile NEWdnr_elev_n82meta. The metadata for the watershed polygons should include citing the report on how those watersheds were delineated and the geoprocessing steps done to classify those watersheds into one of four types. If EPA standards and FGDC compliance cannot be met the contractor shall describe what additional metadata would be required to meet those standards and compliance.

Deliverables: A draft of the metadata that is included with the GIS data for the 82 sample points and watershed polygons that meets EPA standards and is FGDC compliant and can be viewed in ArcGIS 9.3.1.

Due Date: Within six weeks of the start of the contract. The Task Order Contracting Officer (TOCOR) will responds with comments on that draft metadata within two weeks of receipt.

Task 2. Produce a Final Version of the Metadata Meeting EPA Standards and FGDC Compliance

The contractor shall create a final version of the metadata based on comments received from the TOCOR that meet EPA standards and FGDC compliance.

Deliverables: GIS datasets containing metadata of the 82 sample points and watershed polygons that meet EPA standards and FGDC compliance. The contractor shall also send the TOCOR a report that describes how the metadata were validated to meet FGDC and EPA standards.

Due Date: Within fours weeks after the TOCOR has responded with comments on the metadata produced under Task 2.

4. Schedule

Weeks	Action
1 - 6	Task 1 Completed & delivered to TOCOR
7 – 8	TOCOR Responds with comments on Task 1
9 - 14	Task 2 Completed & delivered to TOCOR

References

Detenbeck, N. E., D. Cincotta, J. M. Denver, S. K. Greenlee, A. R. Olsen, and A. M. Pitchford. 2005. Watershed-based survey designs. Environmental Monitoring and Assessment **103**:59-81.

Detenbeck, N. E. and D. A. Cincotta. 2008. Comparability of a regional and state survey: effects on fish IBI assessment for West Virginia, USA. Hydrobiologia **603**:279-300.

Detenbeck, N. E., L. A. Jagger, S. L. Stark, and M. A. Starry. 2004. Watershed Classification Framework for the State of West Virginia: WV R-EMAP Final Report. EPA/600/R-03/141. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Mid-Continent Ecology Division, Duluth, MN.

Table 1 Initial GIS Datasets that produced the 82 watershed polygons and 82 sample points used in the multivariate spatial data analysis

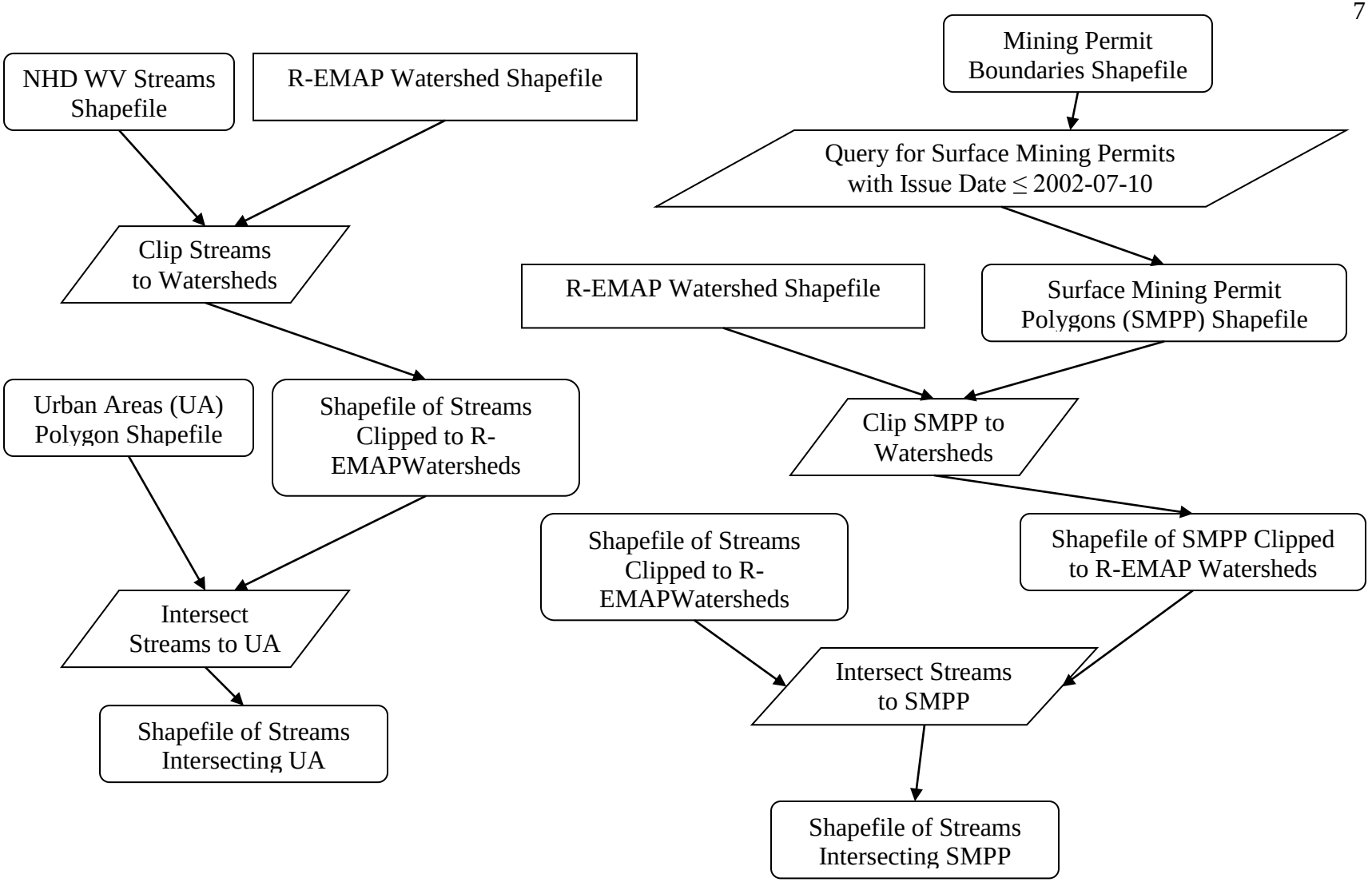
Dataset Name	Description	Format	Source/Point of Contact
Wvrare	Water chemistry, physical habitat, and metrics measured at REMAP sample points	SAS dataset	Lou Reynolds, U.S. EPA Region 3, Wheeling, West Virginia
Remap_dnr	Fish species collected at REMAP sample points, includes geographic coordinates	Excel Spreadsheet	Lou Reynolds, U.S. EPA Region 3, Wheeling West Virginia
Remap01pt_m Remap02pt_m	Shapefile of REMAP sample points	ESRI® point shapefile	Jo Thompson U.S. EPA ORD Mid-Continent Ecology Lab, Duluth MN, and Matthew Starry SRA International, Inc. Contractor to U.S. EPA, ORD, Duluth MN
Remap01_ws_m Remap02_ws_m	Shapefile of REMAP watersheds	ESRI® polygon shapefile	Jo Thompson U.S. EPA ORD Mid-Continent Ecology Lab, Duluth MN, and Matthew Starry SRA International, Inc. Contractor to U.S. EPA, ORD
Mining Permit Boundaries (perbd.zip)	Shapefile of mining permit boundaries	ESRI® polygon shapefile	West Virginia GIS web site http://gis.wvdep.org/data/omr.html

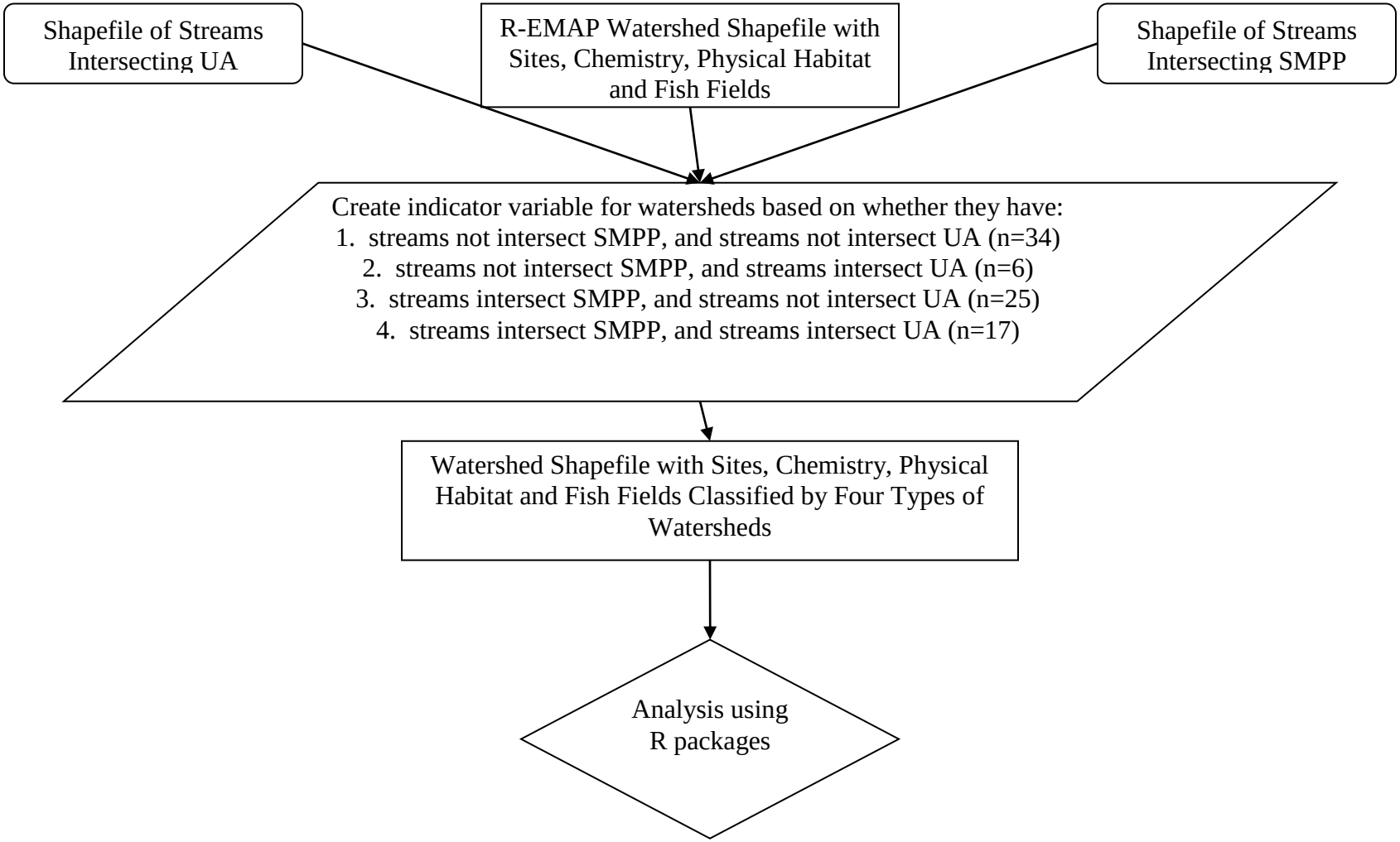
Urbanized Areas (urbanAreas500k_USCenuss_2000_utm83_shp.zip)	Shapefile of urbanized areas	ESRI® polygon shapefile	West Virginia GIS web site http://wvgis.wvu.edu/data/dataset.php?ID=206
National Hydrography Dataset (NHDH)	Shapefile of West Virginia stream network	ESRI® line shapefile	West Virginia GIS web site http://wvgis.wvu.edu/data/dataset.php?ID=235
Level III Ecoregions	Shapefile of Level III Ecoregions	ESRI® polygon shapefile	U.S. EPA web site http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm
West Virginia Boundary	Shapefile of West Virginia	ESRI® polygon shapefile	U.S. EPA GIS WorkGroup EPA Regions Shapefiles

Table 2 Final GIS data sets resulting from geoprocessing of initial data sets in Table 1

Dataset Name	Description	Format	Source/Point of Contact
NEWdnr_elev_n82meta	Shapefile of 82 REMAP sample points used in multivariate spatial data analysis	ESRI® point shapefile	Michael McManus, U.S. EPA, ORD National Center for Environmental Assessment, Cincinnati OH
Ws_n82_eco69_70_v1meta	Shapefile of 82 REMAP watershed polygons used in multivariate spatial data analysis	ESRI® polygon shapefile	Michael McManus, U.S. EPA, ORD National Center for Environmental Assessment, Cincinnati OH

Figure 1. Geoprocessing streams, mining permit boundaries, and urban areas. Rounded-edged rectangles represent initial/intermediate data/GIS layers, parallelograms represent geoprocessing steps, and rectangles represent final data/GIS layers.




Shapefile of Streams Intersecting UA
R-EMAP Watershed Shapefile with Sites, Chemistry, Physical Habitat and Fish Fields
Shapefile of Streams Intersecting SMPP
Create indicator variable for watersheds based on whether they have:
1. streams not intersect SMPP, and streams not intersect UA (n=34)
2. streams not intersect SMPP, and streams intersect UA (n=6)
3. streams intersect SMPP, and streams not intersect UA (n=25)
4. streams intersect SMPP, and streams intersect UA (n=17)
Watershed Shapefile with Sites, Chemistry, Physical Habitat and Fish Fields Classified by Four Types of Watersheds
Analysis using R packages

Appendix A: Metadata on statistical and GIS merger of REMAP water chemistry and fish datasets
Revised: 04/20/2012

I. Water Chemistry Data, which includes physical habitat data, IBI, etc.

A. Original files include: 1. Remap0102pt_m.dbf file obtained from Matthew Starry in Duluth, MN as part of a GIS point shapefile of 122 unique points each with unique HUC_12 variable on 05/12/2010; 2. Wvrare SAS dataset of 128 sites obtained from Lou Reynolds in Wheeling, WV on 05/05/2010 with 118 unique HUC_12s, and the 10 duplicated HUC_12s were because visit_no=2.
B. In SAS merged the two datasets. Used the HUC_12 variable that was in each dataset to do the merge. The merged dataset, remap_rare_ds1, had 127 records, with 117 having visit_no=1, and 10 having visit_no=2, so split that into two datasets: remap_rare_visit1 and remap_rare_visit2.
C. Exported remap_rare_visit1 from SAS as a dbf so the 117 sites and water chemistry at those sites, it could be used in GIS.
D. SAS program: fish_mines.sas does the merger and export.

II. Fish Data

A. Original File obtained from Lou Reynolds as Excel spreadsheet on 01/19/2012: REMAP_DNRdata_concatenate_QC_lou. xls for 124 sites
B. On 04/17/2012 Lou Reynolds said sites DNR_Totals_SITE_ID= "PIGEONCREEK4" and "MIDDLEWHEEL8" should have visit_no be changed from original visit_no=2 to visit_no=1. Visit_no=2 entry in database is wrong for those two sites.
C. In SAS imported the tabs "QA_MAN" and "YOUR NEW FISH DATA" and merged them by site_id and yyyymmdd fields. Split the merged dataset by visit_no=1 and visit_no=2. Dataset fish_site_new_visit1 had 115 sites and fish_site_new_visit2 had 9 sites. From fish_site_new_visit1 dropped DNR_Totals_SITE_ID="MIDDLEFORK13" as Lou said is actively being limed.
D. Exported 114 sites from SAS as dbf file fish_site_new_visit1b to use in GIS.
E. SAS program remap_dnr.sas does this merger and export.

III. Point Shapefile Geoprocessing in GIS

Water Chemistry and Fish Point shapefiles created from dbfs described above and called wq_visit1 and NEWdnr_fish_n114, respectively. Use add XY Data tool and because these are geographic coordinates x field: LON_DD, y field: LAT_DD, choose geographic coordinate system gcs_wgs_1984, datum: d_wgs_1984

A. Spatial join. Analysis Tools > Overlay Toolbox > Spatial Join tool:
   Target: NEWdnr_fish_n114
   Join: wq_visit1
   Output: NEWdnr_fish_spatialJ1

Chose options of join one-to-one, checked keep all target features, match option was closest with a 500 m radius, and record distance between nearest fish site and wq site as fishwq_d. This geoprocessing is taking the water quality data and joining to the fish site and fish data.

B. NEWdnr_fish_n114 had 114 records, but two DNR_Totals sites, "Bakerscree7" and Hurricanecr4", were dropped as those wq sites were not within 500 m of a fish sites so this produced the shapefile: NEWdnr_fish_n112_SpatialJ2 having 112 sites containing fish and chemistry data.

C. Criteria applied to drop 30 sites out of the 112 in NEWdnr_fish_n112_SpatialJ2

1. Other mining permits intersect streams for sites initially classified as ws_type=0, and have high water chemistry values.

a) Deckerscree1: 3rd highest manganese,7th highest specific conductivity, maximum concentration in iron & permit ID E004100 intersects stream
b) Littletenmi1: high calcium and sulfate, and permit ID U200410 intersects stream
c) ~~Pawpawcree5~~ now called Pawpawcree4: second highest calcium and third highest sulfate & permit ID U007883 intersects stream

2. Karst causes inability to measure stream surface network on NHD High resolution so drop 13 sites by selecting subwatersh="Greenbrier".

a) Anthonycree1
b) Easforkgr21
c) Howardcreek1
d) Milligancre1
e) Muddycreek1
f) Muddycreek7
g) Northforkd5
h) Secondcreek24
i) Secondcreek9
j) Sitlingtonc2
k) Springcreek3
l) Westforkgr1
m) Wolfcreek9

3. Fish from different biogeographic area. Drop 7 sites in Potomac watershed

a) Abramscreek 8, which also was a low pH (5.1) site
b) Abramscreek 9, which also has max manganese concentration & SMPP S200409 that parallels stream and within 0.5 km of sampling site .
c) Mikesrun2
d) Newcreek4
e) Northforkp10
f) Pattersoncr11
g) Stonyriver9

4. Norther Panhandle sites with watershed extending outside of WV boundary. Note two wq_visit1 sites, which don't have fish sites, also have watersheds outside boundary.

a) Harmoncreek1
b) Buffalocree38
c) Middlewheel8

5. Ecoregions: Did select by location of select from NEWdnr_wqvisit1_SpatialJ2 that are within Ecoregions69_70_Project_UTM, and that selected 96 out of 110. Reversed the selection shows the 14 sites not in those ecoregions. Ten of those 14 already accounted, and the four that are not are underlined: a) Anthonycree1, b) Eastforkgr21, c) Filescreek7, d) Howardcreek1, e) Laurelfork55, f) Leadingcree17, g) Mikesrun2, h) Newcreek4, i) Northforkd5, j) Northforkp10, k) Pattersoncre11, l) Sitlingtonc2, m) Tygartvalle8, and n) Westforkgr1

D. Wrote query, "drop30sites.exp", applied query, switched selection and resulted in creating the point shapefile, NEWdnr_fish_SpatialJ2_n82.

IV. Watershed Polygon joining to fish & chemistry data

A. Both watershed polygon, ws_n95_eco69_70v2, shapefile and NEWdnr_fish_SpatialJ2_n82 share HUC_12 field. Ws_n95_eco69_70 has ws_type codes for the four watershed types. Ws_n95_eco69_70v2 derived from remap01ws_m that has 122 polygons that I received from Matthew Starry.

B. With watershed polygon selected in Table of Contents, Data Management Tools > Joins Tool > Add Join
Input Join Field: HUC_12
Join Table: NEWdnr_fish_SpatialJ2_n82
Output Join Field: HUC_12
Unchecked Keep All
After ran tool, did data export so created watershed polygon shapefile called: ws_n82_eco69_70_v1, which has fish and chemistry data joined to polygon data. Did remove join so remap01ws_m restored to 122 polygons. The sample break down among the watershed types are:
Ws_type=0 (smpp0ua0) n = 34
Ws_type=1 (smpp0ua1) n = 6
Ws_type=2 (smpp1ua0) n=25
Ws_type=3 (smpp1ua1) n = 17

V. Use of USGS Hydrography Event Management (HEM) Tool

A. Imported sample points so they could be snapped the NHD High resolution geodatabase of NHD flowlines

B. Used Measure Linear Distance and Create Multiple Events Upstream to obtain stream kilometer measurements and total stream kilometers upstream from a sample point

Appendix B: Names of Fish Species used in Multivariate Spatial Data Analysis and Their Concatenated Name of the First Four Letters of the Genus and First Five Letter of the Species

Fish Species	Concatenated Name
1. Campostoma anomalum	CAMPANOMA
2. Cyprinella spiloptera	CYPRSPILO
3. Etheostoma blennioides	ETHEBLENN
4. Etheostoma caeruleum	ETHECAERU
5. Etheostoma flabellare	ETHEFLABE
6. Etheostoma nigrum	ETHENIGRU
7. Etheostoma zonale	ETHEZONAL
8. Hypentelium nigricans	HYPENIGRI
9. Luxilus chrysocephalus	LUXICHRYS
10. Nocomis micropogon	NOCOMICRO
11. Notropis atherinoides	NOTRATHER
12. Notropis buccatus	NOTRBUCCA
13. Notropis photogenis	NOTRPHOTO
14. Notropis rubellus	NOTRRUBEL
15. Notropis stramineus	NOTRSTRAM
16. Notropis telescopus	NOTRTELES
17. Notropis volucellus	NOTRVOLUC
18. Pimephales notatus	PIMENOTAT
19. Semotilus atromaculatus	SEMOATROM